

---

# Research ethics and data sharing

Libby Bishop

UK Data Service – University of Essex

Research Ethics Committee Meeting

University of East London

20 April 2016

---

UK Data Service

---



---

# RECs face competing responsibilities

- Privacy: protecting human subjects, researchers' and organisations' reputations
- and,
- Openness: maximising the value of research data collected from public funding, transparency, research integrity

# Overview

- Big picture – privacy, confidentiality, and consent
- Informed consent
  - Process consent
  - Participants' request to review data – transcriptions, video
  - Withdrawal of consent
- “Alternative legitimate basis” to consent
- Research without consent
  - Covert research and deception
  - Criteria for research without consent
- Public and private – including online
- Vulnerable participants
- Social media
- Open research data
- Data security
- Further guidance for REC members



---

# Sensitive data can be shared - ethically

- Timescapes Qual Longit data
  - sensitive data; 95%+ consent
- Foot and mouth disease in N. Cumbria
  - highly sensitive community information
  - 40/54 interviews; 42/54 diaries; audio restricted
- Bio-bank data—enduring consent
  - no time limits; open reuse
  - 99% consent rate (2500+ patients) – Wales Cancer Bank
- Conflict Archive on the Internet
- Witness to Guantanamo

<http://ukdataservice.ac.uk/manage-data/plan/checklist.aspx>



---

# Privacy, confidentiality, and consent

- Privacy – control over information about oneself (but can also be about physical space)
- Confidentiality – can only pertain to information
- Valid consent
  - Informed
  - Voluntary
  - With capacity



# Data Protection Act, 1998

- Personal data:
  - relate to a living individual
  - individual can be identified
- DPA does not apply to anonymised data
- Only disclose personal data if consent given to do so (and if legally required to do so)
- 5.26 "The DPA is still commonly cited as a reason not to release information when it may be perfectly legitimate to do so." Data Sharing Review
- All change...GDPR 2016

- processed fairly and lawfully
- obtained and processed for specified purpose
- adequate, relevant and not excessive for purpose
- accurate
- not kept longer than necessary
- processed in accordance with the rights of data subjects, e.g. right to be informed about how data will be used, stored, processed, transferred, destroyed; right to access info and data held
- kept secure
- not transferred abroad without adequate protection



---

# Data Protection Act and research

- Exceptions for personal data collected as part of research:
  - can be retained indefinitely (if needed)
  - can be used for other purposes in some circumstances
  - people should still be informed

The Data Protection Act is not intended to, and does not, inhibit ethical research



# Modes of consent

	PROS	CONS
One-off	<ul style="list-style-type: none"><li>• Simple</li><li>• Least hassle to participants</li></ul>	<ul style="list-style-type: none"><li>• Research outputs not known in advance</li><li>• Participants will not know all info they will contribute</li></ul>
Process	<ul style="list-style-type: none"><li>• Ensures 'active' consent</li></ul>	<ul style="list-style-type: none"><li>• May not get all consent needed before losing contact</li><li>• Repetitive, can annoy participants</li></ul>

service



# Form of consent

Written	<ul style="list-style-type: none"><li>• More solid legal ground, e.g. participant has agreed to disclose confidential info</li><li>• Often required by IRB</li><li>• Offers more protection for researcher</li><li>• Not possible for some cases: infirm, illegal activities</li></ul>
Verbal	<ul style="list-style-type: none"><li>• Can be difficult to make all issues clear verbally</li><li>• Possibly greater risks for researcher</li><li>• Best if recorded</li></ul>



# In practice: wording in consent form / information sheet

<b>Use of the information I provide beyond this project</b>		
I agree for the data I provide to be archived at the UK Data Archive. <sup>2</sup>	<input type="checkbox"/>	<input type="checkbox"/>
I understand that other genuine researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.	<input type="checkbox"/>	<input type="checkbox"/>
I understand that other genuine researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.	<input type="checkbox"/>	<input type="checkbox"/>

As the ESRC is a publicly funded body, it has developed ways to share data among academic researchers (subject to strict conditions). To this end, we hope you will allow your anonymised transcript to be stored as part of the UK Data Archive (a service provider for the Economic and Social Data Service).



# Informed consent – information sheets and forms

- Meet requirements of Data Protection laws:
  - purpose of the research
  - what is involved in participation
  - benefits and risks
  - **mechanism of withdrawal**
  - usage of data – for primary research and sharing
  - strategies to ensure confidentiality of data, where relevant
- Complete for all purposes: use, publishing, sharing
- Simple and avoiding excessive warnings



---

# Is consent always necessary?

- Why consent in the first place?
- When is covert research, deception permitted
  - Research has great benefits
  - Risk of harm is low, and has been minimised
  - Research design requires deception
- Case: Correspondence testing for employment discrimination
- But, there is legitimate research without consent
  - Administrative data – not collected for research purposes
  - Public settings
  - Social media – if public?
  - Evaluation research
  - Persons lacking capacity



# Emerging models - consent

- Broad consent – Genome Project, biobanks. Accepted in UK, but not across Europe (e.g., Germany)
- Dynamic consent – maintain contact, seek consent for any new use “substantially” different from original
- “Indeed, European and domestic laws provide several alternatives to consent as the means of legitimising the processing of personal data.” *Data Sharing Review*
  - Public interest, service provision, research
- Alternative legitimate basis (in data protection laws)
  - Administrative Data Research Network
  - Access review committees – medical, Understanding Society
- Proposed EU General Data Protection **Regulation**



---

# Anonymising quantitative data - tips

- remove direct identifiers  
e.g. names, address, institution, photo
- reduce the precision/detail of a variable through aggregation  
e.g. birth year vs. date of birth, occupational categories, area rather than village
- generalise meaning of detailed text variable  
e.g. occupational expertise
- restrict upper lower ranges of a variable to hide outliers  
e.g. income, age
- combining variables  
e.g. creating non-disclosive rural/urban variable from place variables



---

# Anonymising qualitative data

- Remove direct identifiers, or replace with pseudonyms – often not essential research info
- Avoid blanking out
- Identify replacements, e.g. with brackets e.g., [City A]
- Keep anonymisation log of all changes– store separately from data files
- Plan or apply editing at time of transcription
- Avoid over-anonymising –balance anonymisation with the need to preserve data integrity
- Consistency within research team and throughout project.

---

# Audio-visual data

Digital manipulation of audio and image files can remove personal identifiers

*e.g. voice alteration, image blurring (e.g. of faces)*

Labour intensive, expensive, may damage research potential of data

Better alternatives:

- obtain consent to use and share data unaltered for research purposes
- avoid mentioning disclosing information during audio recordings

# In practice: example anonymisation

Ex 1. Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (study 5407 in UK Data Archive collection) by M. Mort, Lancaster University, Institute for Health Research.

Date of Interview: 21/02/02

Interview with **Lucas Roberts**, DEFRA field officer

Date of birth: **2 May** 1965

Gender: Male

Occupation: Frontline worker

Location: **Plumpton**, North Cumbria

**Lucas** was living at home with his parents, "but I'm hoping to move out soon" so we met at his parents' small neat house. We sat in a very comfortable sitting room with an open fire and **Lucas** made me coffee and offered shortbread. Although at first **Lucas** seemed a little nervous, quick to speech and very watchful he seemed to relax as we spoke and to forget about the tape.

**I will just start by asking you to tell me a little bit about yourself and your background.**

Well it is an agricultural background. I grew up on the farm where my brother is now. After I left school I did work on the farm but went to college and did exams, did land use recreation, sort of countryside/ environmental management course. So I obviously left agriculture, did the course and came back [to the farm] at weekends.

Comment [v1]: Replace: Ken

Comment [v2]: delete

Comment [v3]: delete

Comment [v4]: Replace: Ken

Comment [v5]: Replace: Ken

Comment [v6]: Replace: Ken

---

# Qual data – particularly challenging to share

- Strong relationships of trust, commitments to confidentiality
- Participant identity difficult to conceal
  - Audio and visual data
- Research locations potentially identifiable
- Difficult to anonymise data without reducing research value
- Research may investigate illegal activities
  
- But potential benefits of data sharing make it imperative to face these challenges



---

## Open data

- Competing drivers – policy, legal, other
  - Be open and share,
  - RCUK Concordat on Open Data
    - <http://www.rcuk.ac.uk/research/opendata/>
  - Protect privacy
- Is it possible to be open and protect privacy?
  - Yes – with good data management practices, and
  - Open – “to the maximum extent possible”
- Recent challenges for open/sharing data
  - New and novel forms of data (“Big Data”)



# Driving data sharing: UK research funders

## [Research Councils UK Common Principles on Data Policy](#)

- *Publicly funded research data are a public good, produced in the public interest, that should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.*
- in accordance with relevant standards and community best practice
- metadata to make research data discoverable
- **legal, ethical, commercial constraints on release of research data**
- recognition for collecting & analysing data; limited privileged use
- acknowledge sources, intellectual contributions, terms & conditions
- use public funds for management and sharing of research data
- [ESRC-Framework for Research Ethics 2015](#)
- *The research lifecycle also includes knowledge exchange and impact activities, the dissemination process **and the archiving, future use, sharing and linking of data***



# Driving data sharing: journals & publishers

- data underpinning publication accessible
  - upon request from author
  - as supplement with publication
  - in public repository

## Examples:

- BioMed Central [open data statement](#)
- Data journals: [Journal of open psychology data](#)
- Psychological Sciences: ‘open data badge’
- PLOS ONE: “Publication is conditional upon the agreement of the authors to make freely available any materials and information described in their publication that may be reasonably requested by others.” .../.... “will not consider a study if the conclusions depend solely on the analysis of proprietary data” ... “the paper must include an analysis of public data that validates the conclusions so others can reproduce the analysis and build on the findings.”



# Driving data sharing: Intl'l research funders

European open access policies: Horizon 2020, European Research Council (ERC)

- [communication & recommendation on access to / preservation of scientific information](#) (July 2012) (publications & research data)
- [pilot on open access to research data](#), primarily data underlying (open access) scientific publications for Horizon 2020
- [data management guidelines](#) for Horizon 2020 (~ policies)

National Science Foundation  
requires data management plans



UK Data Service

generally based on [OECD Principles and Guidelines for Access to Research Data from Public Funding](#)



# Ethical arguments for sharing/open

- Duties to scholarly community
  - Transparency
  - Research integrity
- Duties to public
  - Use public funds well
- Duties to participants – protect **and**
  - Avoid burdensome replication
  - Optimise use of their data



---

# Even when data cannot be made open...

- Not all data can be made open, if we mean open = public
- Publish/advertise:
  - Which data exist
  - Where data are kept, e.g. which repository
  - Who can access them
  - For what purposes
  - Under which conditions



---

# Public/private

- There are often clear(ish) guidelines for each domain, problem is the boundaries are blurring
- This is bigger issue than just physical/virtual space
  - Public/private sector
  - Definitions of “public interest” and “public good”
- But one big problem area is online research
  - Some say – if area is theoretical public, then public rules apply, no need for consent
  - But, surveys show many people expect privacy in these areas, so others are seeking consent even in “public” spaces
  - Some are defining a new category, open+private, available to anyone in the public, but have to act to enter



# Vulnerability

- What does it mean to be vulnerable?
  - Lack of capacity – unable to protect own interests
  - At risk of (exceptional harm)
  - But...this could apply to many kinds of participants...
- Usually, vulnerability implies additional factors
  - Limited capacity
  - Voluntariness compromised in some way (duress, unequal power)
  - Physical limitations
- Generally agreed that to assess vulnerability, must look at
  - Individual characteristics, but also
  - Context, situation, nature of decision, degree of risk
- Finally, need to balance duty to protect the vulnerable with duty to enable equitable participation (not exclude on the basis of vulnerability)



# Internet research – social media, etc.

- Association of Internet Researchers <http://aoir.org/ethics/>
- British Psychological Soc'ty – internet mediated research  
<http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf>
- COSMOS – Cardiff Online Social Media Observatory Ethics Resource Guide & Statement
  - <https://www.cs.cf.ac.uk/cosmos/ethics-resource-guide/>
  - <https://www.cs.cf.ac.uk/cosmos/cosmos-ethics-statement/>
- New social media, new social science (#NSMNSS)
  - <http://nsmnss.blogspot.co.uk/>
- Wisdom of the Crowd – IPSOS Mori, ESRC,, et al.  
<https://www.ipsos-mori.com/ourexpertise/digitalresearch/sociallistening/wisdomofthecrowd/publications.aspx>



# COSMOS Lab Ethics Framework

- We follow the ESRC's Framework for Research Ethics
- All projects undergo Research Committee Review
- Any significant changes to research design following are reported back to the Committee for re-approval
- We abide by Twitter's Developer Policy and Developer Agreement
- In most cases we only publish in research outputs aggregate information based on data derived from the Twitter APIs
- **In research outputs we never directly quote identifiable individual Twitter users without informed consent.** Where informed consent cannot be obtained we represent the content of tweets in aggregate form (e.g. topic clustering, wordclouds) and themes (decontextualised examples and descriptions of the meaning or tone of tweet content). These forms of representation preclude the identification of individual Twitter users, preserving anonymity and confidentiality
- **In research outputs we do directly quote from identifiable Twitter accounts maintained by public and commercial organisations (e.g. government departments, law enforcement, local authorities, companies) without seeking prior informed consent**
- **We never share data gathered from Twitter APIs for our research outside of the Lab or project partners**
- **\*\*\*And they have procedures for gaining consent for individual tweets\*\***

---

# Twitter, tweets and consent (1 of 2)

- Public tweets may appear similar to public space, but there are legal and morally significant factors to consider
- Constraints from Twitter's Terms and Conditions
  - Tweets may not be anonymised
  - Authors must honour requests for change, deletion
    - Impossible if paper with quotes has been published
- Researcher has duty of care to tweeter
  - Whether content has been made public, or not



# Twitter, tweets and consent (2 of 2)

- For small scale data projects, consent is possible
  - Contact tweeter directly to request consent
  - Full information can be provided via a website
  - Must make clear to tweeters they will NOT be anonymous
    - Even if tweetID is not real name, tweet text can be searched and re-linked with tweeter
- When consent is not possible, there may be alternatives
  - Data “bricolage” (Markham 2012)  
<http://www.tandfonline.com/doi/abs/10.1080/1369118X.2011.641993>
- In selected cases, consent may not be needed
  - Aggregated data – no identification possible
  - Tweets from public, official or institutional accounts
  - See COSMOS Risk Assessment

<http://the-sra.org.uk/wp-content/uploads/ethics-in-social-media-research-matthew-williams.pdf>

With thanks to Matthew Williams & Luke Sloan (Cardiff University)

UK Data Service



# Big data with privacy issues

- Some big data analytics will involve processing personal data:
  - social media
  - loyalty cards
  - sensors in clinical trials
- Recent ICO report: Big data and data protection

<b>Personal data</b>	Does your big data project need to use personal data at all? If you are using personal data, can it be anonymised? If you are processing personal data you have to comply with the Data Protection Act.
<b>Privacy impact assessments</b>	Carry out a privacy impact assessment to understand how the processing will affect the people concerned. Are you using personal data to identify general trends or to make decisions that affect individuals?
<b>Repurposing data</b>	If you are repurposing data, consider whether the new purpose is incompatible with the original purpose, in data protection terms, and whether you need to get consent. If you are buying in personal data from elsewhere, you need to practice due diligence and ensure that you have a data protection condition for your processing.
<b>Data minimisation</b>	Big data analytics is not an excuse for stockpiling data or keeping it longer than you need for your business purposes, just in case it might be useful. Long term uses must be articulated or justifiable, even if all the detail of the future use is not known.
<b>Transparency</b>	Be as transparent and open as possible about what you are doing. Explain the purposes, implications and benefits of the analytics. Think of innovative and effective ways to convey this to the people concerned.
<b>Subject access</b>	People have a right to see the data you are processing about them. Design systems that make it easy for you to collate this information. Think about enabling people to access their data on line in a re-usable format.



---

# References and resources

- Barocas, S. and Nissenbaum, H. (2014) Big Data's End Run around Anonymity and Consent, in J. Lane et al. (eds) *Privacy, Big Data and the Public Good*. Cambridge University Press.
- Chessell, M. (2014) Ethics of Big Data and Analytics, [http://www.ibmbigdatahub.com/sites/default/files/whitepapers\\_reports\\_file/TCG%20Study%20Report%20-%20Ethics%20for%20BD%26A.pdf](http://www.ibmbigdatahub.com/sites/default/files/whitepapers_reports_file/TCG%20Study%20Report%20-%20Ethics%20for%20BD%26A.pdf).
- Narayanan, A. and Felton, E. (2014) No silver bullet: de-identification still doesn't work, <http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf>.
- ICO. Big data and data protection. <https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf>
- ESRC case studies - vulnerability, deception, and more
- <http://www.esrc.ac.uk/funding/guidance-for-applicants/research-ethics/ethics-case-studies/>



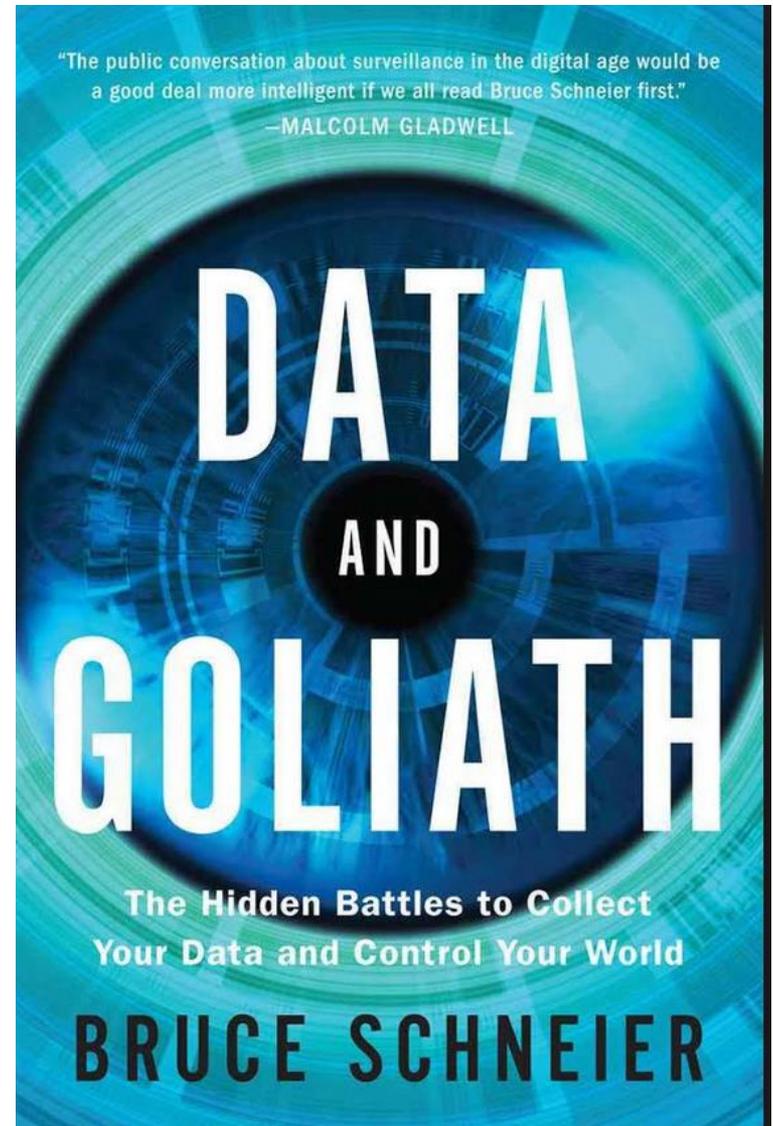
# References and resources (cont.)

- The Research Ethics Guidebook (Boddy, Morrow)
- <http://www.ethicsguidebook.ac.uk/>
- European textbook on ethics in research
- [https://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/textbook-on-ethics-report\\_en.pdf](https://ec.europa.eu/research/science-society/document_library/pdf_06/textbook-on-ethics-report_en.pdf)
- A quick overview of changes to applying for ethics with NHS:  
<http://www.hra.nhs.uk/research-community/hra-approval-the-new-process-for-the-nhs-in-england/>
- There is also a (quite complex looking) flowchart of the process for ethical approval: <http://www.hra.nhs.uk/documents/2014/05/rec-application-process-flowchart-v6-0-19-may-2014.pdf>



...how do we design systems that make use of our data collectively to benefit society as a whole, while at the same time protecting people individually?...This is it: this is the fundamental issue of the information age.”

Bruce Schneier 2015 *Data and Goliath*



---

# Contact

Collections Development and Producer Relations team  
UK Data Service  
University of Essex  
[ukdataservice.ac.uk/help/get-in-touch.aspx](http://ukdataservice.ac.uk/help/get-in-touch.aspx)

UK Data Service



---

# Questions

Libby Bishop

ebishop@essex.ac.uk

Collections Development and Producer Relations

UK Data Service

University of Essex

[ukdataservice.ac.uk/help/get-in-touch.aspx](http://ukdataservice.ac.uk/help/get-in-touch.aspx)

